

p53 Mutations Associated with Breast, Colorectal, Liver, Lung, and Ovarian Cancers

Tamar Lasky and Ellen Silbergeld

Department of Epidemiology and Preventive Medicine, University of Maryland at Baltimore, Baltimore, MD 21201-1596 USA

In this paper we describe a statistical analysis of the European Molecular Library *p53* mutation database comparing *p53* mutations occurring in breast, colorectal, liver, lung, and ovarian cancers. The analyses show that mutation hot spots vary by cancer and that base pair changes and predicted amino acid changes in the gene product vary by cancer and by codon. The analyses use relative frequencies and epidemiologic measures of effect (prevalence ratios) not applied previously to these data. The five cancers in the database with the largest sample sizes were breast (418), colorectal (398), liver (341), non-small cell lung (313), and ovarian cancers (251), for a total of 1,721 reports of *p53* mutations. The five cancers varied considerably in the distribution of mutations over sites, with different hot spots in each cancer. At the six most frequently reported codon sites, we compared base pair and amino acid changes by type of cancer. The comparison of base pair changes indicated a predominance of particular base pair changes at a codon (for example, C→T and G→A changes at Codon 248) and their association with specific cancers (C→T changes with colorectal cancer and G→A changes with both colorectal and breast cancers at codon 248). Comparing predicted amino acid changes by codon and cancer was also intriguing, as in codons 175 and 273, where arginine to cysteine and arginine to histidine changes were frequent in breast, colorectal, and ovarian cancers. Variations in *p53* mutational distributions by cancer may be explained by different exposures to carcinogens or by organ-specific clonal selection. Further research may be stimulated by this analysis. *Key words:* amino acid, cancer, database, epidemiology, gene, molecular epidemiology, mutation, *p53*. *Environ Health Perspect* 104:1324–1331 (1996)

The nature and significance of the distribution of mutations by base pair in the *p53* gene have been the subject of investigation in recent years because of the importance of this gene in control of the cell cycle, DNA repair and synthesis, cell differentiation, and cell death (1–9). Mutations in the *p53* gene have been found in many types of cancer and may be the most common mutation observed in human cancers. In lung cancer, *p53* mutations have been found in 56% of tissue samples, and in colorectal, esophageal, ovarian, pancreatic, and skin cancers, prevalences of 44–50% have been reported (4). Mutations occur at many base pairs throughout the gene, with higher frequencies reported in exons 5–8 than at other exons and high frequencies occurring at specific codons (4,10–12). The diversity in *p53* mutations has suggested the possibility of correlating exposures with specific *p53* mutations, for example, vinyl chloride and mutations at A:T base pairs in angiosarcomas, ultraviolet radiation and CC→TT changes in skin cancer, aflatoxins and G→T mutations in liver cancer, or G:C to A:T transitions associated with mustard gas exposure in lung cancer patients (13–19). Such epidemiologic correlations may be supported by laboratory studies showing specific mutations associated with specific mutagens. For example, *n*-ethyl-*n*-nitrosourea, an alkylating agent associated

with carcinogenesis in the gut and other tissues, induces G→A transitions at codon 248 in human fibroblasts (20).

These observations suggest that in cancer cells derived from tumors in different organs, *p53* mutational spectra can be seen as the fingerprints of the carcinogens that caused the cancer (10,11,19,21–23). The hypothesis that environmental carcinogens correlate strongly with *p53* mutational spectra is the basis for suggestions that certain *p53* mutations can be interpreted as markers of exposure. Some examples of these suggestions are 1) that the presence of *p53* mutations in esophageal carcinoma patients with and without human papilloma virus involvement is evidence of the role of environmental carcinogens in esophageal carcinogenesis (24); 2) that varying frequencies of microdeletions, transitions, and transversions in different breast cancer populations are compatible with the role of different environmental carcinogens contributing to breast cancer carcinogenesis in different populations (25); 3) that differences in proportions of transversions in Japanese and American prostate cancer patients suggests variation in etiologic factors between the two populations (26); and 4) that observations of the presence of G→T changes in breast cancer and those observed in lung cancer (which the authors maintain are probably

caused by exogenous mutagenic chemicals) suggest that these breast cancers were similarly caused by exogenous chemicals (23).

Researchers have characterized mutational variation primarily by summarizing aggregate base pair changes throughout the gene and presenting the data in pie charts, histograms, or tables. These analyses indicate variation in base pair changes by cancer: for example, differences in the proportions of A:T to C:G changes in cancers of the skin, nasopharynx, oral cavity, and pharynx/larynx, or differences in mutation type (deletion, base pair substitution, splice site, etc.) (4,21,27). Other studies have analyzed DNA change by location; these have reported variations in hot spots (frequent codon sites) in bladder, prostate, breast, and other cancers (28–30). However, most of the analyses have used individual data sets, limited by small sample sizes that do not permit statistical tests. The availability of the pooled database compiled by the European Molecular Biology Library has allowed analyses of larger sample sizes, but these primarily have been summaries of base pair changes by cancer (mutational spectra) (4,31,32).

To fully evaluate the assumption that *p53* mutational spectra reflect specific environmental carcinogens, it is necessary to fully describe and characterize variation in *p53* mutational spectra; the second step is to correlate variation in *p53* mutational spectra with specific exposures. In this paper, we extend the evaluation of *p53* mutational variation by analyzing an internationally available database of *p53* mutations. We expand on previous analyses by characterizing differences in frequent codon sites in five cancers—breast, colorectal, liver, lung, and ovarian—and by examining differences by cancer in base pair changes at six codon sites that have been identified in the literature as *p53* mutation hot spots.

Address correspondence to T. Lasky, Department of Epidemiology and Preventive Medicine, University of Maryland at Baltimore, Gray Lab, Room 209, 520 West Lombard Street, Baltimore, MD 21201-1627 USA.

We thank Curt Harris, Ih-Chang Hsu, and Judith Stamberg for their comments. The analysis was supported by a grant from the Heinz Family Foundation.

Received 26 July 1996; accepted 10 September 1996.

We also analyzed the predicted amino acid change at codon sites by cancer and tested for statistical significance by comparing relative frequencies. Because the database did not contain information on exposures, we were not able to correlate mutational variation with exposures.

Methods

The European Molecular Biology Library has compiled reports from individual studies to produce an aggregated database of *p53* mutations (31–33). The version used in these analyses contained data describing 4,123 mutations and became available over the Internet on 29 September 1995 (<http://sunsite.unc.edu/dnam/mainpage.html>). For each mutation reported, the database described the cancer diagnosis and type, the base pair change, the type of mutation (base pair substitution, frameshift, deletion, insertion, splicing mutation, complex mutation), base pair position, codon position, wild type codon, codon produced by mutation, wild type amino acid, amino acid produced by mutation, and references. The database included reports from 443 articles published from 1989 to 1995. Sixty-three of the articles contained only one report of a *p53* mutation, and 296 of the articles reported fewer than 10 mutations. Only 12 articles reported more than 40 mutations.

The data were derived both from studies that analyzed specific exons (5–8 only), as well as from studies that sequenced the entire gene. Thus, the database may over-represent mutations in exons 5–8, but information about other exons sequenced was not available in the database. When the same mutations were reported in more than one journal article, only one report was entered in the database, but methods for identifying duplicate reports were not explicitly stated. Although the database contains information regarding the type of cancer, diagnosis and terminology were adopted as used by the authors of the original journal articles. Thus, “esophageal” and “oesophageal” were entered as two different cancers in the database, an indication that terminology in journal articles was not standardized when entering data. Despite its limitations, the database offered a unique opportunity to analyze adequate sample sizes of reports of *p53* mutations and to explore new approaches to analyzing such databases.

Frequency counts by type of cancer, codon site, and amino acid change were obtained using the SAS 6.10 statistical package (SAS, Cary, NC). We identified the five most frequent codon sites in each cancer to consider variation in hot spots by type of cancer. To compare variation in

mutations at codon sites, we chose six codon sites that have been noted as hot spots in the literature and analyzed mutations in the five cancers (breast, colorectal, liver, lung, and ovarian) with the largest sample sizes. Three-dimensional histograms of base pair changes by cancer and amino acid changes by cancer at specific codons revealed distinct nonrandom patterns that varied by cancer and by codon. The associations between amino acid changes and cancer type were then assessed using prevalence ratios, an epidemiologic measure of effect that compares the prevalence (the proportion of the group with a characteristic) in one group to the prevalence of the same characteristic in a second group (34). A prevalence ratio greater than 1 indicates that the event is more frequent in the first group than in the second group; a prevalence ratio of 3, for example, indicates a prevalence three times greater in the first group than in the second group. The 95% confidence interval (CI) for each prevalence ratio was calculated: a CI that overlaps 1 means that the estimate of the

prevalence ratio may include 1 (suggesting no difference between the two groups). *p*-Values were calculated using Fisher's exact chi-square.

Results

Mutation hot spots by cancer. The five cancers in the database with the largest sample sizes were breast (418), colorectal (398), liver (341), non-small cell lung (313), and ovarian (251), for a total of 1,721 reports of *p53* mutations. The five cancers varied considerably in the distribution of mutations over sites (Table 1). The five most frequent sites for *p53* mutations differed. In breast cancer, they were (in order) 273, 175, 248, 245, and 249. For liver cancer, the five most frequent sites were codons 249, 273, 166, 157, and 159. The cancers also varied in the proportion of mutations occurring at these codon sites. While codon 249 was a frequent site in both breast and liver cancers, only 2.2% of *p53* mutations in breast cancer occurred at codon 249, compared to 36.4% of the mutations in liver cancer. Colorectal and

Table 1. The five most frequent codon sites (hot spots) for *p53* mutations in breast, colorectal, liver, lung, and ovarian cancers and the percentages of *p53* mutations occurring at these sites for each cancer

Codon (hot spot)	Breast (n = 418)	Colorectal (n = 398)	Liver (n = 341)	Lung (n = 313)	Ovarian (n = 251)
Most frequent (%)	273 (6.7%)	248 (13.8%)	249 (36.4%)	273 (8.9%)	273 (11.2%)
Second most frequent (%)	175 (6.0%)	175 (13.6%)	273 (4.1%)	248 (4.5%)	248 (5.6%)
Third most frequent (%)	248 (6.0%)	273 (9.5%)	166 (3.5%)	179 (4.2%)	282 (5.2%)
Fourth most frequent (%)	245 (3.1%)	245 (7.0%)	157 (1.8%)	245 (4.2%)	245 (4.8%)
Fifth most frequent (%)	249 (2.2%)	282 (5.0%)	159 (1.8%)	157 (3.5%)	175 (4.0%)
Cumulative percentage of five most frequent codon sites out of total gene	24.0%	48.9%	47.6%	25.3%	30.8%

Table 2. Most frequent base pair change at codons in five cancers

Codon (hot spot)	Breast (n = 418)	Colorectal (n = 398)	Liver (n = 341)	Lung (n = 313)	Ovarian (n = 251)
Most frequent	273 G→A (19/28)	248 C→T (31/55)	249 G→T (118/124)	273 G→T (11/28)	273 G→A (14/28)
Second most frequent	175 G→A (22/25)	175 G→A (49/54)	273 C→T (9/14)	248 G→T (5/14)	248 C→T (7/14)
Third most frequent	248 G→A (16/25)	273 G→A (21/38)	166 T→A (12/12)	179 A→G (8/13)	282 C→T (10/13)
Fourth most frequent	245 G→A (8/13)	245 G→A (23/28)	157 G→T (3/6)	245 G→T (7/13)	245 G→A (9/12)
Fifth most frequent	249 G→T (4/9)	282 C→T (20/20)	159 G→C (5/6)	157 G→T (10/11)	175 G→A (10/10)
Most frequent base pair change reported for entire gene	G→A (28.0%; 117/418)	G→A (37.2%; 148/398)	G→T (42.5%; 145/341)	G→T (30.7%; 96/313)	G→A (30.3%; 76/251)

ovarian cancers shared the same five hot spots, but the relative frequencies differed. In colorectal cancer, 13.8% of the mutations occurred at codon 248 (which was the most frequent site), but in ovarian cancer, only 5.6% of the mutations occurred at that codon (the second most frequent site). Lung cancer shared two hot spots with breast, colorectal, and ovarian cancers (codons 273 and 248). Overall, 10 different codons were identified as hot spots in one or more of the five cancers.

Table 2 shows the most frequent base pair change reported at the five most frequent sites for each cancer and the most frequent base pair changes reported throughout the gene for each cancer. In breast, colorectal, and ovarian cancers, G→A changes were reported in 28.0–37.2% of all *p53* mutations, and G→A mutations were frequently reported at hot spots. C→T changes were frequent at two hot spots in colorectal cancer and in ovarian cancer, and G→T changes were frequent in the fifth hot spot in breast cancer. In both liver and lung cancers, G→T base pair changes predominated (42.5% and 30.7% of all reported mutations, respectively), but the two cancers shared only two hot spots, codons 157 and 273. At the five hot spots in liver cancer, the most frequent base pair changes were G→T, C→T, T→A, G→T, and G→C (from most frequent site to fifth most frequent site), but in lung cancer, G→T base pair changes predominated at four of the five hot spots.

Thus, the distribution of base pair changes over hot spots varied by cancer. While G→A base pair changes were the most frequently reported changes in all *p53* mutations in breast, colorectal, and ovarian cancers, G→A changes predominated at four hot spots in breast cancer and G→A and C→T were frequent at hot spots in both colorectal and ovarian cancers. Similarly, while G→T changes were the most frequently reported changes in all *p53* mutations in liver and lung cancers, G→T changes predominated at four hot spots in lung cancer, but four different base pair changes predominated at hot spots in liver cancer.

Base pair and amino acid changes by cancer for six codons. The next analysis examined the diversity of mutations at six *p53* mutation hot spots identified in the literature, codons 248, 245, 249, 175, 273, and 282, by type of cancer, to see if specific or distinctive mutations were associated with the cancers. Because different cancers do not appear to share similar hot spots, as shown in the previous analysis, we could not compare the types of mutations at the five most frequent mutation sites in each

cancer. Instead we compared base pair and amino acid changes by cancer at six hot spots identified in the literature (4). In addition to describing the most frequent base pair changes at each codon by cancer, we compared relative frequencies by codon and by cancer to find mutations associated with each of the five cancers.

Table 3 summarizes the number of reported *p53* mutations at codons 248, 245, 249, 175, 283, and 282 and their relative frequencies by type of cancer and specific amino acid changes that resulted. Thus, there were nine reports of arginine (Arg) to tryptophan (Trp) changes at codon 248 among 418 reports of all *p53* mutations in breast cancer, for a relative frequency of 2.15%, and 31 such mutations out of 398 *p53* mutations reported in colorectal cancer, for a relative frequency of 7.78%. As noted previously in Table 1, few codons are the sites for more than 10% of the mutations reported for a type of cancer. For all five cancers, the most frequent site accounted for 6.7–36.4% of all *p53* mutations, and the fifth most frequent site accounted for only 1.8–5.0% of reported *p53* mutations. Analyzing individual types of mutations at a codon site describes events that account for even lower percentages of the total. For example, while codon

273 was the most frequent site of *p53* mutations in breast cancer (6.7%), the Arg to histidine (His) changes at codon 273 comprised 19, or 4.54%, of all breast cancer *p53* mutations, the Arg to cysteine (Cys) changes accounted for 6, or 1.43%, and the Arg to leucine (Leu) change accounted for 1, or 0.23%.

To estimate the strength of association between specific mutations and cancers and to test the statistical significance of the variation, the relative frequency of particular amino acid changes at a given codon in one type of cancer was compared to the relative frequency of the same change in the other cancers (Table 4). For example, Arg to Trp changes at codon 248 accounted for 7.79% of all the *p53* mutations reported in colorectal cancer, but only 1.59% of the *p53* muta-

Table 4. Calculation of prevalence ratios

Cancer	Codon 248, Arg to Trp	All other mutations at any codon	Total
Colorectal cancer	31 (7.79%)	367	398
Breast, liver, lung, and ovarian cancers	21 (1.59%)	1,302	1,323

Prevalence ratio (colorectal vs. breast, liver, lung, and ovarian) = 4.9; 95% CI = 2.9–8.4; $p < 0.001$.

Table 3. Number and percentages of specific base pair and amino acid changes out of the total *p53* mutations reported for each of five cancers

Codon	Amino acid and base pair change	Breast	Colorectal	Liver	Lung	Ovarian
248 (CGG)	Arg to Trp	9	31	1	4	7
	C→T	(2.15%)	(7.78%)	(0.29%)	(1.27%)	(2.79%)
	Arg to Gln	16	23	4	3	5
	G→A	(3.83%)	(5.78%)	(1.17%)	(0.95%)	(1.99%)
	Arg to Leu	0	1	0	6	0
245 (GGC)	G→T		(0.25%)		(1.92%)	
	Gly to Ser	4	14	0	2	8
	G→A	(0.95%)	(3.52%)		(0.63%)	(3.19%)
	Gly to Asp	4	9	0	3	1
	G→A	(0.95%)	(2.26%)		(0.95%)	(0.40%)
249 (AGG)	Gly to Cys	1	2	3	5	0
	G→T	(0.23%)	(0.50%)	(0.87%)	(1.60%)	
	Arg to Ser	5	2	114	8	2
	G→T	(1.19%)	(0.50%)	(33.43%)	(2.55%)	(0.80%)
	Arg to His	22	49	2	2	10
175 (CGC)	G→A	(5.26%)	(12.31%)	(0.58%)	(0.63%)	(3.98%)
	Arg to His	19	21	1	5	14
	G→A	(4.55%)	(5.28%)	(0.29%)	(1.59%)	(5.5%)
	Arg to Cys	6	17	9	3	13
	C→T	(1.43%)	(4.27%)	(2.63%)	(0.95%)	(5.18%)
273 (CGT)	Arg to Leu	1	0	3	11	0
	G→T	(0.23%)		(0.87%)	(3.51%)	
	Arg to Trp	3	20	2	8	10
	C→T	(0.71%)	(5.02%)	(0.58%)	(2.55%)	(3.98%)
	Total mutations analyzed above	90 (21.53%)	189 (47.49%)	139 (40.76%)	60 (19.17%)	70 (27.89%)
Total mutations reported		418	398	339	313	251

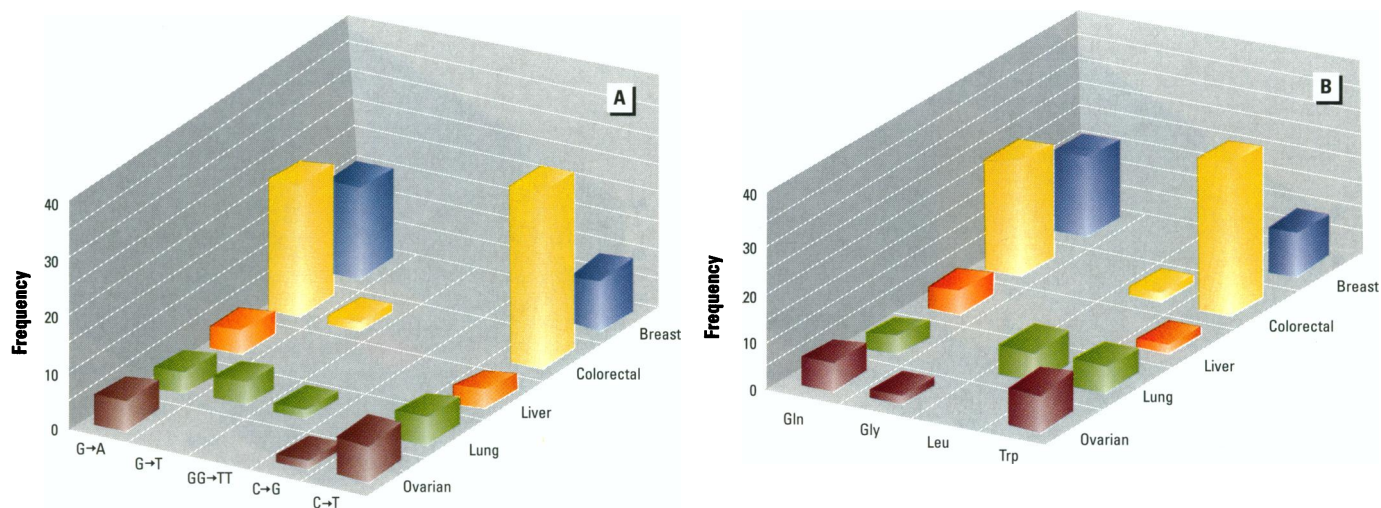


Figure 1. Frequency of changes at codon 248 for breast, colorectal, liver, lung, and ovarian cancers. (A) Frequency of base pair changes (CGG); GG→TT indicates that two base pair changes were observed in the same report. (B) Frequency of amino acid changes (Arg).

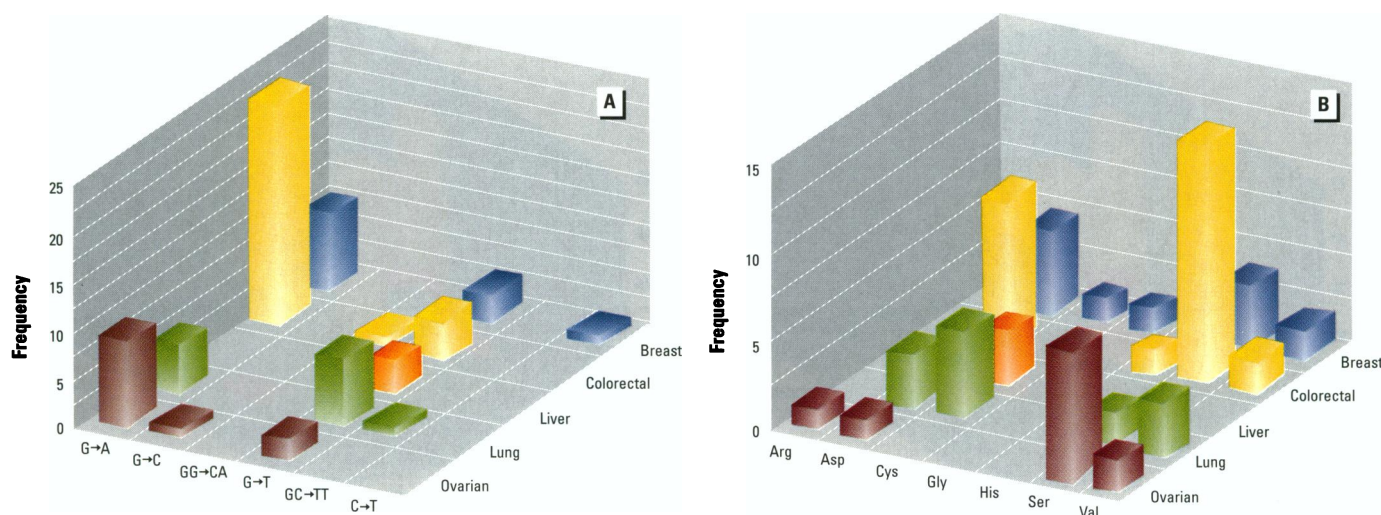


Figure 2. Frequency of changes at codon 245 for breast, colorectal, liver, lung, and ovarian cancers. (A) Frequency of base pair changes (GGC); GG→CA and GC→TT indicate that two base pair changes were observed in the same report. (B) Frequency of amino acid changes (Gly).

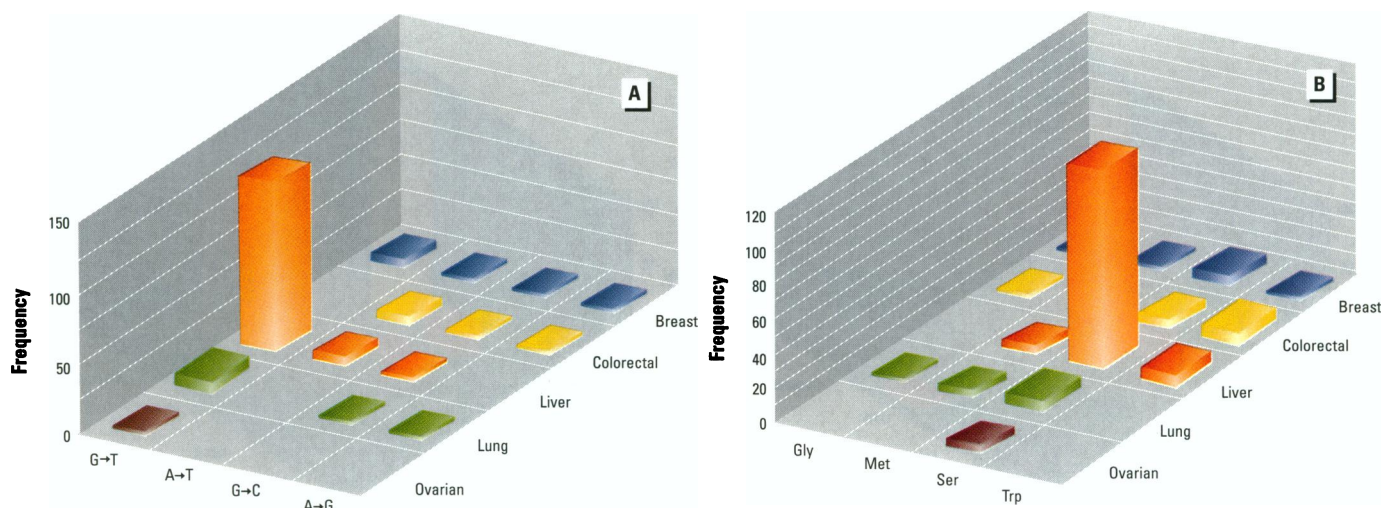


Figure 3. Frequency of changes at codon 249 for breast, colorectal, liver, lung, and ovarian cancers. (A) Frequency of base pair changes (AGG). (B) Frequency of amino acid changes (Arg).

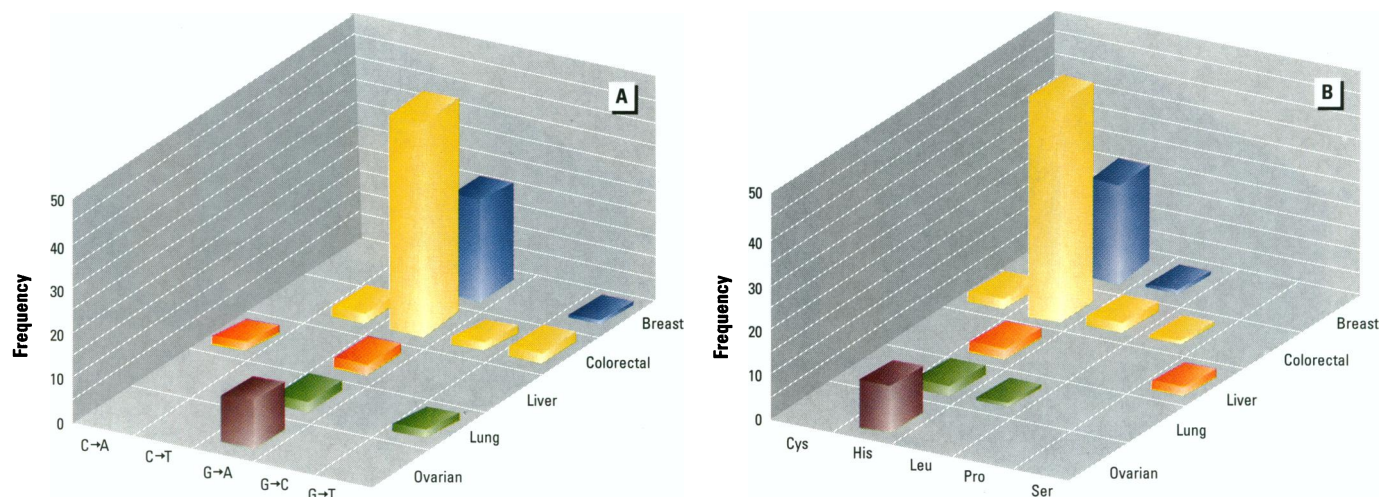


Figure 4. Frequency of changes at codon 175 for breast, colorectal, liver, lung, and ovarian cancers. (A) Frequency of base pair changes (CGC). (B) Frequency of amino acid changes (Arg).

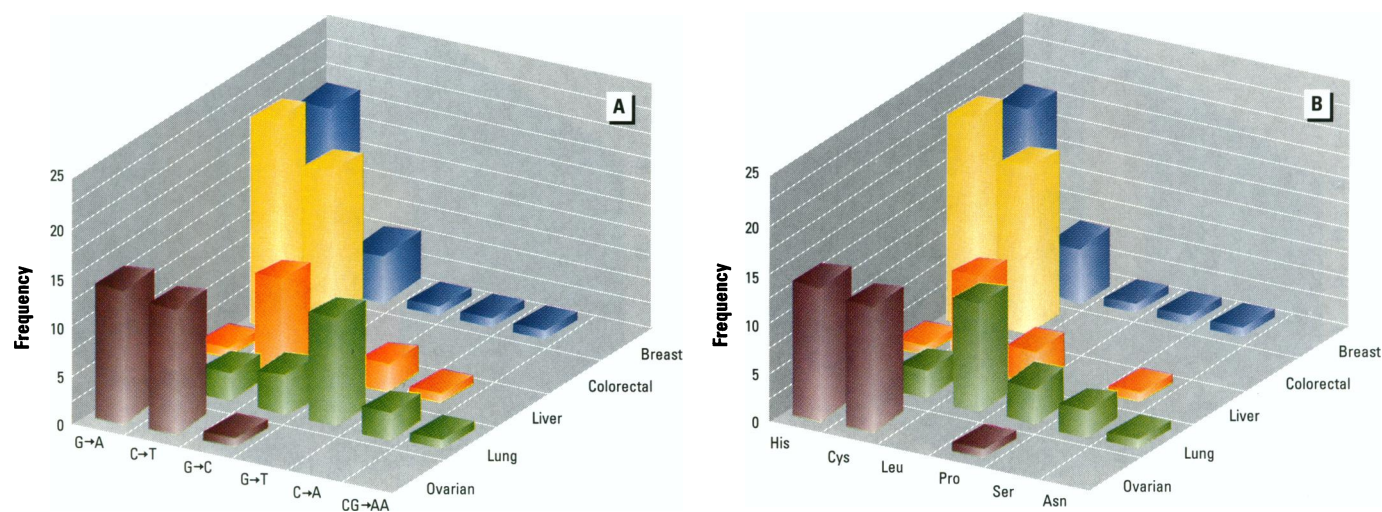


Figure 5. Frequency of changes at codon 273 for breast, colorectal, liver, lung, and ovarian cancers. (A) Frequency of base pair changes (CGT); CG→AA indicates that two base pair changes were observed in the same report. (B) Frequency of amino acid changes (Arg).

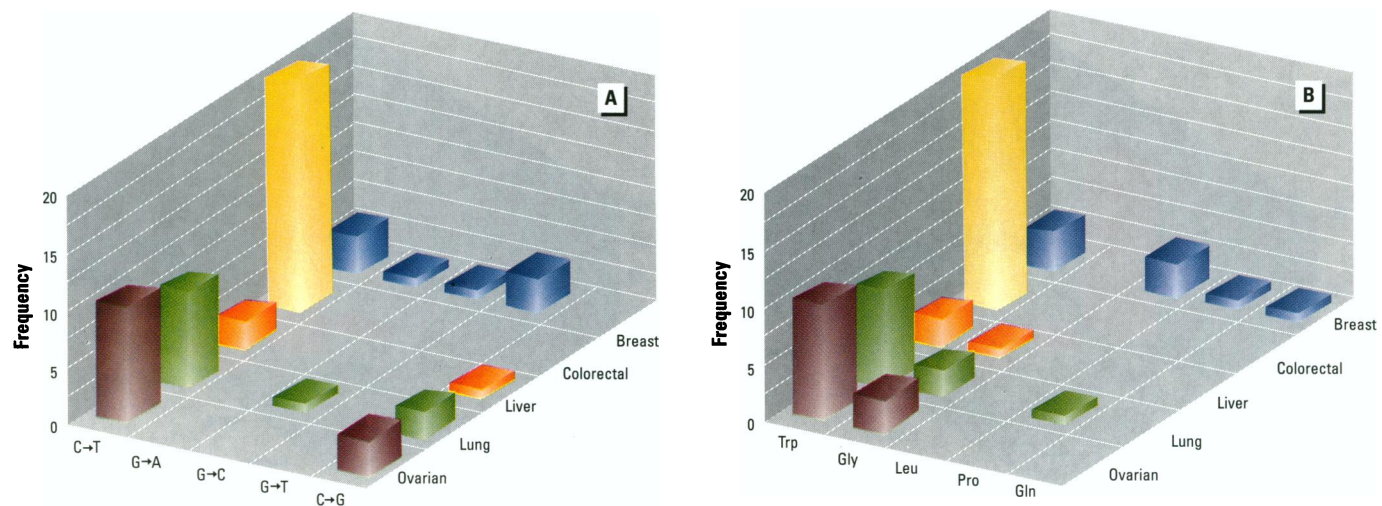


Figure 6. Frequency of changes at codon 282 for breast, colorectal, liver, lung, and ovarian cancers. (A) Frequency of base pair changes (CGG). (B) Frequency of amino acid changes (Arg).

tions reported in breast, liver, lung, and ovarian cancers. Thus, it was 4.9 times more likely to be observed in colorectal cancer than in the other four cancers. The 95% CI of this ratio was 2.9–8.4 with a p -value <0.001 , indicating that the difference is greater than that expected by chance alone. We did not develop a statistical method for determining appropriate comparisons, but relied on histograms to suggest comparisons.

Three-dimensional histograms of mutation frequency by base pair change and cancer or amino acid change and cancer were useful in visualizing the distinct patterns at each of the codons analyzed (Figs. 1–6). Figure 4A and B, for example, show the frequency distributions of mutations at codon 175 by type of base pair change and by amino acid change. The histograms indicate that G→A (Arg to His) changes may be associated with ovarian, colorectal, and

breast cancers, but not with liver or lung cancers. Figure 5A and B show the more complex pattern of mutations occurring at codon 273. Six different base pair changes occurred, resulting in six different amino acid changes; however, most of the amino acid changes are to His or Cys except for lung cancer, where Arg to Leu (G→T) changes appeared more frequently. In colorectal cancer, Arg to His and Arg to Cys (G→A and C→T) were the only changes reported at this codon.

The three-dimensional histograms were used to suggest comparisons for statistical analysis, and Table 5 describes the prevalence ratios for specific amino acid changes at codons 245, 248, 249, 175, 273, and 282. For example, at codon 248, the prevalence of Arg to Trp changes in colorectal cancer was compared to the prevalence of those events in the other four cancers. At the same codon, the prevalence of Arg to glutamine (Gln)

changes in breast cancer was compared to its prevalence in liver, lung, and ovarian cancers; the prevalence of the same changes in colorectal cancer was also compared to the prevalence in liver, lung, and ovarian cancers. At codon 248, Arg to Trp (C→T) changes were associated with colorectal cancer only (compared to the other four cancers), but Arg to Gln (G→A) changes were associated with both breast and colorectal cancers compared to the other three cancers. Arg to Leu (G→T) changes were associated with lung cancer only. At codon 245, glycine (Gly) to serine (Ser) (G→A) changes were associated with colorectal and ovarian cancers, Gly to aspartic acid (Asp) (G→A) was associated with colorectal cancer only, and Gly to Cys (G→T) was associated with lung cancer only. At codon 249, Arg to Ser (G→T and G→C) changes were strongly associated with liver cancer only. At codon 175, Arg to His (G→A) changes were associated with breast and liver cancers, but not with colorectal, lung, or ovarian cancers. At codon 273, Arg to His (G→A) changes were associated with breast, colorectal, and ovarian cancers, Arg to Cys (C→T) changes were associated with colorectal and ovarian cancers, and Arg to Leu (G→T) changes were associated with lung cancer only. At codon 282, Arg to Trp (C→T) changes were associated with colorectal, lung, and ovarian cancers, but not with breast or liver cancers. All of the above associations were statistically significant.

Discussion

The five most frequent sites of *p53* mutations were different in breast, colorectal, lung, liver, and ovarian cancers. This is consistent with previous reports in the literature that hot spots and mutational spectra vary by cancer (1,4,12,29,32). The variation in hot spots is not explained by the predominance of specific base pair changes in each cancer. Thus, while G→T base pair changes were the most frequent changes in both liver and lung cancers, the two cancers shared only two hot spots. The great majority (118/145) of G→T changes in liver cancer occurred at codon 249, its number one hot spot, but codon 249 was not a hot spot in lung cancer. Even at a shared hot spot, codon 273, different patterns of base pair changes were observed, with C→T more frequent in liver cancer and G→T more frequent in lung cancer. In some of the cancers, the percentages of mutations were low even at the most frequent sites, suggesting the need to further define the term hot spot and also suggesting that some cancers may not have true hot spots for *p53* mutations.

In our analyses we compared relative frequencies as well as absolute numbers of events. These statistical methods identified

Table 5. Prevalence ratios, 95% confidence intervals, and p -values for associations between amino acid changes and cancer at six codons

Codon	Amino acid and base pair change	Breast	Colorectal	Liver	Lung	Ovarian
248 (CGG)	Arg to Trp C→T	—	4.9 ^a (2.9–8.4) $p<0.001$	—	—	—
	Arg to Gln G→A	2.9 ^b (1.4–6.0) $p<0.006$	4.4 ^b (2.2–8.7) $p<0.001$	—	—	—
	Arg to Leu G→T	—	—	—	27.0 ^c (3.3–223.4) $p<0.001$	—
245 (GGC)	Gly to Ser G→A	—	6.3 ^d (2.4–16.2) $p<0.001$	—	—	5.7 ^d (2.0–16.3) $p<0.002$
	Gly to Asp G→A	—	3.7 ^a (1.5–9.6) $p<0.007$	—	—	—
	Gly to Cys G→T	—	—	—	3.7 ^c (1.15–12.2) $p<0.05$	—
249 (AGG)	Arg to Ser G→T (G→C)	—	—	27.6 ^e (16.8–45.3) $p<0.001$	—	—
175 (CGC)	Arg to His G→A	3.4 ^b (1.8–6.6) $p<0.001$	8.0 ^b (4.4–14.2) $p<0.001$	—	—	—
273 (CGT)	Arg to His G→A	5.0 ^f (2.0–12.3) $p<0.001$	5.8 ^f (2.3–14.1) $p<0.001$	—	—	6.1 ^f (2.4–15.6) $p<0.001$
	Arg to Cys C→T	—	2.5 ^d (1.3–4.9) $p<0.007$	—	—	3.1 ^d (1.5–6.2) $p<0.003$
	Arg to Leu G→T	—	—	—	12.4 ^c (4.0–38.6) $p<0.001$	—
282 (CGT)	Arg to Trp C→T	—	7.6 ^g (2.9–20.2) $p<0.001$	—	3.9 ^g (1.3–11.8) $p<0.03$	6.0 ^g (2.1–17.5) $p<0.001$

^aColorectal cancer compared to the other four cancers.

^bBreast or colorectal cancer compared to the other three cancers.

^cLung cancer compared to the other four cancers.

^dColorectal or ovarian cancer compared to the other three cancers.

^eLiver cancer compared to the other four cancers.

^fBreast, colorectal, or ovarian cancers compared to liver and lung cancers.

^gColorectal, lung, or ovarian cancers compared to breast and liver cancers.

associations even when absolute frequencies were low, as in the example of ovarian cancer and *p53* mutations at codon 282. Only 13 mutations were reported at codon 282 in patients with ovarian cancer, but 10 of these were C→T mutations resulting in Arg to Trp changes in the codon. The prevalence of these Arg to Trp changes in the sample of mutations reported in ovarian cancer was 3.98%, six times the prevalence observed in breast and liver cancers. In this example, the small number of mutations at the codon consisted primarily of one particular amino acid change and the clustering of one type of change was not explained by chance variation. At some sites, more than one type of base pair or amino acid change was associated with some cancers but not with others. Both C→T and G→A base pair changes were associated with colorectal cancer at codons 248 and 273; only G→A changes were associated with breast cancer at these codons; and both C→T and G→A changes were associated with ovarian cancer only at codon 273. It is possible that the colorectal and ovarian cancer tissue samples used in studies in the database consisted of different groups: one exposed to carcinogens causing G→A changes and one exposed to carcinogens causing C→T changes, but the data do not permit such an analysis. The unique distribution of G→A and C→T changes at codons 248 and 273 in breast and ovarian cancers is not explained by the fact that these are CpG sites.

These analyses point out the value of describing the distribution of mutations throughout the gene instead of reporting aggregate summaries of the total base pair changes. Additionally, information on resultant amino acid changes provides further information. Such analysis allowed us to consider the diversity of mutations at specific codons by type of cancer and by their possible effects on the gene product. Changes to Cys at codon 245, His at 175, and His or Cys at 273 suggest that some mutations may have a greater effect than others on the structure and function of the *p53* gene product.

Two factors have been postulated to explain variation in the distribution of *p53* mutations by cancer (4,29). The first is the role of different exposures and their effect on the *p53* gene. If different mutagens have distinct effects on *p53* mutational spectra and are associated with cancers in different tissues, one would expect to observe mutation distributions that vary by cancer type. A second source of variation may be clonal selection factors that could modify the originally induced mutational spectra to those that are eventually observed in tumor cells. If clonal selection factors vary by organ, then

cancers in different organs would show different distributions of *p53* mutations and such variation would also correlate with the different carcinogens associated with different cancers.

If the carcinogens are the main determinants of *p53* mutation distributions, similar patterns should be observed in patients with similar carcinogen exposures, even if their cancer occurs in different organs. Thus, lung cancer and bladder cancer in smokers would be expected to show similar *p53* mutation distributions. If organ-specific clonal selection forces are the main determinants of *p53* mutation distributions, variation by cancer type should be observed regardless of exposure etiology. To test these hypotheses, the association between the *p53* mutation distribution and the cancer type must be evaluated with adequate patient sample sizes, appropriate statistical tests, and confirmed exposure histories. For example, estimation of the independent effects of radon and cigarette smoking on *p53* mutations would require samples from tumors collected from four groups: radon-exposed patients who were nonsmokers, nonsmokers without radon exposure, patients exposed to both, and patients exposed to neither smoking nor radon. As Esteve et al. (21) have pointed out, access to tissue samples with detailed exposure histories presents its own set of methodologic issues, usually relying on paraffin-embedded tissue that may be sub-optimal for genetic analysis.

Lacking validated exposure histories, we cannot test these hypotheses with the database we analyzed. It is, of course, possible that both exposure and clonal selection shape the mutational spectra that are eventually observed, as suggested by the patterns of mutations occurring at codon 273. Codon 273 is the only codon that is a hot spot for all five cancers, and G→A and C→T base pair changes at codon 273 frequently result in amino acid changes to Cys or His in breast, colorectal, and ovarian cancers, both of which might be expected to alter the conformation and function of the *p53* protein. Both types of base pair changes are associated with colorectal and ovarian cancers, but G→A (Arg to His) changes are associated with breast cancer only. In lung cancer, mutations at codon 273 are most frequently Arg to Leu changes resulting from G→T base pair changes. The pattern observed in lung cancer is consistent with suggestions that G→T transversions are more common in smokers than in nonsmokers and are frequently observed in small-cell lung cancer and that benzo[*a*]pyrene diol epoxide (BPDE) adducts form at codon 273(4,35).

Inferences drawn from these data are limited by the data quality and by the lack of crucial information, particularly complete sequencing of the *p53* gene and ascertainment of exposure histories. In breast cancer, 22% of *p53* mutations have been reported to lie outside exons 5–8, although all missense mutations were reported within exons 5–8; thus, analyses relying on studies with differences in the exons sequenced may report different results (36). Inferences are also limited by the sample, a nonrandom, nonrepresentative selection of patients that may not truly reflect the distribution of *p53* mutations in these cancers. The database is also limited by the absence of data describing age, gender, ethnicity, exposure history, severity of illness, co-morbidity, medications, and other characteristics of the patients. It was not possible to assess whether this was a representative sample of patients or which of these patient characteristics could have affected the overall distribution of mutations observed in the database.

The variation in *p53* mutation distributions is more complex than has been previously described. The distributions vary by cancer in the sites (hot spots) of frequent mutations, the relative frequencies of mutations at hot spots, and specific base pair and amino acid changes at individual codons. Specific amino acid changes are associated with different cancers, and these patterns were statistically significant. Variations in organ-specific exposures and variations in organ-specific clonal selection factors may explain these associations, but a definitive assessment of the contribution of these forces to variation, as well as description of other correlates of *p53* mutation distributions, awaits further epidemiologic and laboratory research.

A great deal of attention has been focused on the relationship between mutagens and observed *p53* mutations without adequately considering the effects of clonal selection and without adequately considering the combined effects of different mutagens on *p53* mutational distributions. Thus, while mutagens may cause distinct *p53* mutations, it is not clear how such patterns might be shaped by exposure to multiple mutagens or by clonal selection forces. It is also not clear whether the correlation between mutagen and *p53* mutational distributions will be consistently observed in molecular epidemiologic studies after modification by clonal selection. There is much work to be done in characterizing the extent and nature of variation in *p53* mutation distributions by cancer. Before describing the extent of such variation, it seems premature to conclude that the vari-

ation is a marker for or evidence of the exposure that caused the cancer and that no other factors influence the distribution of *p53* mutations.

REFERENCES

- Harris CC, Hollstein M. Clinical implications of the *p53* tumor-suppressor gene. *N Engl J Med* 329:1318–1327 (1993).
- Selivanova G, Wiman KG. *p53*: a cell cycle regulator activated by DNA damage. *Adv Cancer Res* 66:143–148 (1995).
- Hollstein M, Sidransky D, Harris CC. *p53* mutations in human cancer. *Science* 253:49–53 (1991).
- Greenblatt MS, Bennett WP, Hollstein M, Harris CC. Mutations in the *p53* tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res* 54:4855–4878 (1994).
- Bogler O, Huang HJ, Kleihues P, Cavenee WK. The *p53* gene and its role in human brain tumors. *Glia* 15:308–327 (1995).
- Kastan MB, Canman CE, Leonard CJ. *p53*, cell cycle control and apoptosis: implications for cancer. *Cancer Metastasis Rev* 14:3–15 (1995).
- Lee JM, Bernstein A. Apoptosis, cancer and the *p53* tumour suppressor gene. *Cancer Metastasis Rev* 14:149–161 (1995).
- Gotz C, Montenarh M. *p53*: DNA damage, DNA repair, and apoptosis. *Rev Physiol Biochem Pharmacol* 127:65–95 (1996).
- Ozbun MA, Butel JS. Tumor suppressor *p53* mutations and breast cancer: a critical analysis. *Adv Cancer Res* 66:71–141 (1995).
- Harris CC. The 1995 Walter Hubert Lecture—molecular epidemiology of human cancer: insights from the mutational analysis of the *p53* tumour-suppressor gene. *Br J Cancer* 73:261–269 (1996).
- Harris CC. 1995 Deichmann Lecture—*p53* tumor suppressor gene: at the crossroads of molecular carcinogenesis, molecular epidemiology and cancer risk assessment. *Toxicol Lett* 82–83:1–7 (1995).
- Harris CC. *p53*: At the crossroads of molecular carcinogenesis and risk assessment. *Science* 262:1980–1981 (1993).
- Aguilar F, Hussain SP, Cerutti P. Aflatoxin B₁ induces the transversion of G to T in codon 249 of the *p53* tumor suppressor gene in human hepatocytes. *Proc Natl Acad Sci USA* 90:8586–8590 (1993).
- Hollstein M, Marion MJ, Lehman T, Welsh J, Harris CC, Martel-Planche G, Kusters I, Montesano R. *p53* mutations at A:T base pairs in angiosarcomas of vinyl chloride-exposed factory workers. *Carcinogenesis* 15:1–3 (1994).
- Takeshima Y, Inai K, Bennett WP, Metcalf RA, Welsh JA, Yonehara S, Hayashi Y, Fujihara M, Yamakido M, Akiyama M. *p53* mutations in lung cancers from Japanese mustard gas workers. *Carcinogenesis* 15:2075–2079 (1994).
- Basset-Seguín N, Moles JP, Mils V, Dereure O, Guilhou JJ. *TP53* tumor suppressor gene and skin carcinogenesis. *J Invest Dermatol* 103:102S–106S (1994).
- Nakazawa H, English D, Randell PL, Nakazawa K, Martel N, Armstrong BK, Yamasaki H. UV and skin cancer: specific *p53* gene mutation in normal skin as a biologically relevant exposure measurement. *Proc Natl Acad Sci USA* 91:360–364 (1994).
- Taguchi M, Watanabe S, Yashima K, Murakami Y, Sekiya T, Ikeda S. Aberrations of the tumor suppressor *p53* gene and *p53* protein in solar keratosis in human skin. *J Invest Dermatol* 103:500–533 (1994).
- Daya-Grosjean L, Dumaz N, Sarasin A. The specificity of *p53* mutation spectra in sunlight induced human cancers. *J Photochem Photobiol B* 28:115–24 (1995).
- Hussain SP, Aguilar F, Cerutti P. Mutagenesis of codon 248 of the human *p53* tumor suppressor gene by *N*-ethyl-*N*-nitrosourea. *Oncogene* 9:13–18 (1994).
- Esteve A, Sorlie T, Martel-Planche G, Hollstein M, Kusters I, Lewalter J, Vineis P, Stephan-Odenthal M, Montesano R. Screening for *p53* gene mutations in archived tumors of workers occupationally exposed to carcinogens: examples from analysis of bladder tumors. *J Occup Environ Med* 37:59–68 (1995).
- Vogelstein B, Kinzler KW. Carcinogens leave fingerprints. *Nature* 355:209–210 (1992).
- Biggs PJ, Warren W, Venitt S, Stratton MR. Does a genotoxic carcinogen contribute to human breast cancer? The value of mutational spectra in unravelling the aetiology of cancer. *Mutagenesis* 8:275–283 (1993).
- Chang F, Syrjänen S, Tervahauta A, Kurvinen K, Wang L, Syrjänen K. Frequent mutations of *p53* gene in oesophageal squamous cell carcinomas with and without human papillomavirus (HPV) involvement suggest the dominant role of environmental carcinogens in oesophageal carcinogenesis. *Br J Cancer* 70:346–351 (1994).
- Hartmann A, Rosanelli G, Blaszyk H, Cunningham JM, McGovern RM, Schroeder JJ, Schaid DJ, Kovach JS, Sommer SS. Novel pattern of *p53* mutation in breast cancers from Austrian women. *J Clin Invest* 95:686–689 (1995).
- Watanabe M, Ushijima T, Kakiuchi H, Shiraishi T, Yatani R, Shimazaki J, Kotake T, Sugimura T, Nagao M. *p53* gene mutations in human prostate cancers in Japan: different mutation spectra between Japan and Western countries. *Jpn J Cancer Res* 85:904–910 (1994).
- Redston MS, Caldas C, Seymour AB, Hruban RH, da Costa L, Yeo CJ, Kern SE. *p53* mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. *Cancer Res* 54:3025–3033 (1994).
- Chi SG, deVere-White RW, Meyers FJ, Siders DB, Lee F, Gumerlock PH. *p53* in prostate cancer: frequent expressed transition mutations. *J Natl Cancer Inst* 86:926–933 (1994).
- Levine AJ, Wu MC, Chang A, Silver A, Attiyeh EF, Lin J, Epstein CB. The spectrum of mutations at the *p53* locus. Evidence for tissue-specific mutagenesis, selection of mutant alleles, and a “gain of function” phenotype. *Ann NY Acad Sci* 768:111–128 (1995).
- Shibata A, Ohneseit PF, Tsai YC, Spruck CR, Nichols PW, Chiang HS, Lai MK, Jones PA. Mutational spectrum in the *p53* gene in bladder tumors from the endemic area of black foot disease in Taiwan. *Carcinogenesis* 15:1085–1087 (1994).
- Cariello NF, Cui L, Beroud C, Soussi T. Database and software for the analysis of mutations in the human *p53* gene. *Cancer Res* 54:4454–4460 (1994).
- Hollstein M, Rice K, Greenblatt MS, Soussi T, Fuchs R, Sorlie T, Hovig E, Smith-Sorensen B, Montesano R, Harris CC. Database of *p53* gene somatic mutations in human tumors and cell lines. *Nucleic Acids Res* 22:3551–3555 (1994).
- Cariello NF, Piegorsch WW, Adams WT, Skopek TR. Computer program for the analysis of mutational spectra: application to *p53* mutations. *Carcinogenesis* 15:2281–2285 (1994).
- Hennekens CH, Buring JE. *Epidemiology in medicine*. Boston:Little, Brown and Company, 1987.
- Denisenco MF, Pao A, Tang M, Pfeifer GP. Preferential formation of benzo[*a*]pyrene adducts at lung cancer mutational hotspots in *p53*. *Science* 274:430–432 (1996).
- Hartmann A, Blaszyk H, McGovern RM, Schroeder JJ, Cunningham J, De Vries EM, Kovach JS, Sommer SS. *p53* gene mutations inside and outside of exons 5–8: the patterns differ in breast and other cancers. *Oncogene* 10:681–688 (1995).

Nutrition and Immunity

May 5–7, 1997 Atlanta, Georgia

Sponsored by
American Cancer Society
Centers for Disease Control and Prevention
Emory University
JLSS Research Foundation
International Conference Series on Nutrition and Health Promotion

For more information, contact:
The Conference on Nutrition and Immunity
International Life Sciences Institute
1126 Sixteenth Street, NW, Washington, DC 20036-4810
(202) 659-0074 Fax: (202) 659-3859 E-mail: meetings@dc.ils.org